



Synthèse de mouvements humains par des méthodes basées apprentissage : un état de l'art

Pamela Carreno-Medrano, Sylvie Gibet, Pierre-François Marteau

► To cite this version:

Pamela Carreno-Medrano, Sylvie Gibet, Pierre-François Marteau. Synthèse de mouvements humains par des méthodes basées apprentissage : un état de l'art. AFIG 2013, Nov 2013, Limoges, France. hal-00913056

HAL Id: hal-00913056

<https://hal.science/hal-00913056>

Submitted on 3 Dec 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Synthèse de mouvements humains par des méthodes basées apprentissage : un état de l'art

Pamela Carreno¹, Sylvie Gibet¹, Pierre-François Marteau¹

¹ Laboratoire IRISA, Université de Bretagne Sud
Campus de Tohannic, Bât. ENSIbs, 56017 Vannes Cedex, France

Résumé

En raison du nombre croissant d'applications graphiques, la génération de personnages se comportant comme des humains est devenue un sujet de recherche important. Différentes approches ont été proposées, celles combinant les données de capture de mouvement et des méthodes d'apprentissage automatique étant la tendance dominante des dernières années. Malgré les bons résultats obtenus par ces approches, à notre connaissance, il n'existe pas à l'heure actuelle de document de synthèse sur les méthodes basées apprentissage qui présentent le problème de la réutilisation et de la généralisation des données de capture de mouvement. Dans cet article, nous présentons un état de l'art des méthodes les plus couramment utilisées pour la synthèse de mouvements cinématiques du corps complet. Nous présentons les principes et les idées sous-jacentes à chacune d'elles, ainsi que le type de représentation et les étapes de pré-traitement appliqué sur les données de capture de mouvement avant apprentissage.

Due to the increasing number of graphical applications, the generation of human life-like characters has become an important research topic. Different approaches have been proposed, the combination of motion capture data and machine learning methods being the dominant trend in the last years. Despite the good results produced by these approaches, to our knowledge there does not exist a document surveying the learning-based methods that intend to solve the problem of re-using and generalizing motion capture data. In this article, we present a state-of-the-art of the most recurrent methods used for the synthesis of whole-body kinematic motions. We present the principles and ideas behind each one of them, as well as the type of representations and pre-processing steps applied over motion capture data before learning.

Mots clé : animation basée données, apprentissage artificiel, synthèse du mouvement humain.

1. Introduction

Avec le développement d'applications numériques mettant en jeu des personnages virtuels, l'un des défis majeurs de l'animation par ordinateur consiste à proposer des méthodes et des logiciels capables de générer des comportements de ces personnages inspirés du monde réel. Un certain nombre d'applications de ce type ont pour utilisateur final l'humain, qui est facilement enclin à percevoir les erreurs les plus subtiles dans les mouvements produits par l'homme. Par conséquent, afin de préserver l'engagement de l'utilisateur dans l'application virtuelle, il est primordial de s'intéresser, non pas uniquement au calcul d'animations convenables, mais de synthétiser des mouvements réalistes et naturels.

L'un des meilleurs moyens de parvenir à cet objectif est d'enregistrer les mouvements d'acteurs réels. Cependant la

capture de données de haute qualité est un procédé coûteux en temps et en ressources matérielles. Dans ces conditions, il est souhaitable de réutiliser les données disponibles[†] plutôt que de capturer de nouveaux mouvements à chaque développement d'une nouvelle application. Or, la modification des données dans l'objectif de les ré-exploiter dans différents contextes s'avère être une tâche difficile.

Plusieurs approches classiques cherchent à résoudre ce problème : *i. Traitement du signal* : en faisant l'hypothèse que le mouvement est un signal échantillonné au cours du temps, cette approche propose d'utiliser des techniques des domaines de l'image et du traitement du signal pour éditer et adapter les données capturées à de nouveaux scénarios [BW95]. *ii. Mélange de mouvement / interpolation* : cette approche permet de produire de nouveaux mouve-

[†]. Dans cet article la notion de réutilisation de données de mouvements capturés signifie l'adaptation à de nouvelles situations, de nouveaux personnages, ou à de nouveaux styles ou expressivités.

ments à partir d'une combinaison linéaire de deux ou plusieurs segments de mouvement similaires. Les poids peuvent être sélectionnés manuellement, calculés automatiquement en fonction de certaines contraintes, ou définis par une fonction mathématique, par exemple à l'aide de fonctions radiales de base (*RBF*). Nous renvoyons le lecteur à [FHKS12] pour une liste détaillée de ces méthodes et la comparaison des techniques qui suivent ce principe. *iii. Concaténation de mouvement* : cette approche propose des structures et des méthodes de ré-assemblage de segments de mouvement afin de produire de longs enchaînement d'actions. [KG*02], [AF02] et [AFO03] sont des exemples de cette dernière catégorie.

Bien que les approches pré-citées réutilisent les données de capture du mouvement et ainsi permettent de produire de nouvelles séquences qui ont le réalisme et les nuances du mouvement original, elle ne permettent pas de produire des mouvements en dehors du spectre des données capturées. Par exemple, pour synthétiser le mouvement d'une personne applaudissant, sautant, ou applaudissant en sautant, il est nécessaire de disposer dans la base de données de tous les mouvements des différentes actions ; à défaut, il n'est pas possible de produire des mouvements satisfaisant de telles contraintes [CH05]. Ceci est dû au fait que ces méthodes ne contrôlent pas totalement le processus de synthèse, ne modifient pas le contenu, n'exploitent pas toute la connaissance ni la structure sous-jacente propre aux données, et ne se généralisent pas à d'autres situations, acteurs ou styles.

Une solution séduisante à ces limitations est de combiner les données de capture de mouvements humains avec des techniques d'apprentissage automatique. Ce faisant, il est possible de bénéficier de toutes les connaissances et l'expérience d'un domaine de recherche dont le but est d'étudier les données et de construire des modèles qui généralisent bien au-delà des exemples. Grâce aux principes et aux techniques du *machine learning*, il est possible de modéliser le processus qui a produit les mouvements enregistrés et d'utiliser ce modèle pour générer de nouvelles données. Celles-ci sont conformes aux exemples, aux attentes et aux contraintes de l'utilisateur, ainsi qu'au contexte dans lequel elles seront utilisées [Her03]. En outre, les méthodes de *machine learning* permettent de capturer toutes les subtilités du mouvement humain et de générer d'autres mouvements experts tout en conservant le style, l'expressivité et la sémantique inhérents aux actions humaines.

Même si les techniques de *machine learning* ont montré qu'elles produisent de bons résultats et sont de plus en plus utilisées par la communauté de l'animation par ordinateur, à notre connaissance, il n'existe pas d'état de l'art des méthodes basées apprentissage qui permettent de résoudre le problème de la réutilisation et de la généralisation des données de mouvement. Plusieurs tours d'horizon ont été proposés [AI06], [PP10] et [VWVBE*10], présentant soit des méthodes basées données "pures" dans lesquelles aucun modèle n'est appris à partir des données, soit des modifications des approches précédentes, soit des méthodes qui réalisent un compromis entre le contrôle et le réalisme (lors de l'utilisation des données de mouvement).

Nous présentons ici un état de l'art des méthodes basées

apprentissage utilisées dans le cadre de l'animation de personnage. Dans cette étude, nous nous focalisons sur des animations du corps entier, car une projection visuelle du corps humain (sans le visage et les mains) est suffisante pour transmettre l'idée d'un mouvement plausible et naturel [Joh73].

Au-delà des travaux récents proposant un couplage entre les méthodes d'animation basées sur la physique, les données de capture de mouvement et les modèles d'apprentissage automatique, cette étude se concentre plus spécifiquement sur l'animation basée cinématique, c'est à dire utilisant les positions et les rotations ainsi que leurs dérivées respectives du premier et du second ordre. Nous adoptons cette approche car les modèles physiques restent difficiles à contrôler et ne garantissent pas que le mouvement résultant soit naturel. En outre, nous considérons que les méthodes utilisées dans un contexte cinématique peuvent être facilement extrapolées aux modèles d'animation basés sur la physique comme l'a montré [WMC11].[‡]

Dans la ligne du travail de synthèse proposé par [MHK06] sur le problème de la capture et la modélisation du mouvement humain basées-vision, nous nous concentrons dans cet article sur une vue d'ensemble et une description des concepts sous-jacents aux techniques d'apprentissage utilisées pour résoudre le problème posé, plutôt que de décrire de manière exhaustive les travaux individuels qui s'y rapportent. En adoptant ce point de vue, nous visons fournir aux lecteurs intéressés une introduction la plus complète possible de l'animation de personnage virtuel au moyen de méthodes d'apprentissage automatique, ainsi qu'une source de références pour les chercheurs déjà experts sur le sujet.

L'article est organisé de la manière suivante : après une discussion sur les différentes représentations du mouvement et les étapes de pré-traitement classiquement opérées sur les données, nous décrivons les principales démarches et méthodes relatives à l'apprentissage automatique dans la section 3. Une discussion et une comparaison entre les méthodes est ensuite proposée dans la section 4. Enfin nous répertorions certaines applications dans la section 5 avant de conclure.

2. Préliminaires

Préalablement à la présentation des méthodes d'apprentissage automatique utilisées pour l'animation de personnages virtuels, nous décrivons brièvement ici quelques éléments de spécification, de représentation et de pré-traitement qui sont partagés par la plupart des approches.

2.1. Spécification du mouvement

D'un point de vue cinématique, un personnage virtuel est représenté par un corps articulé constitué d'une structure arborescente composée de segments rigides (os) connectés par des articulations. Afin de modifier la posture du personnage,

[‡]. Les lecteurs intéressés par l'animation basée physique peuvent se référer à [GP12]

des transformations sont appliquées sur toutes les articulations de la hiérarchie du squelette. Ainsi, la configuration du personnage (posture x_i) est définie par :

$$x_i = \{(p_i^1, q_i^1), (p_i^2, q_i^2), \dots, (p_i^k, q_i^k)\} \quad (1)$$

où (p_i^1, q_i^1) sont respectivement les position et orientation du noeud racine, (p_i^j, q_i^j) sont les position et orientation relatives associées à la $j^{\text{ème}}$ articulation, avec $2 \leq j \leq k$ et k le nombre total d'articulations dans la hiérarchie. Chaque posture x_i correspond à un vecteur de dimension $6 * k$.

Un mouvement peut être représenté à partir d'une séquence de postures $X = \{x_1, x_2, \dots, x_t\}$ de longueur t et de dimension $D = 6 * k * t$, échantillonnées régulièrement dans le temps.

2.2. Représentation du mouvement

Images clés ou postures : c'est sans doute la représentation la plus utilisée par les modèles d'apprentissage. Les mouvements sont représentés par des vecteurs de postures de grande dimension, chaque posture étant elle-même un vecteur de positions, rotations, ou les deux. Afin de préserver l'information temporelle du mouvement, des vecteurs de vitesse et d'accélération cartésienne ou angulaire peuvent être ajoutés aux vecteurs de postures. On définit ainsi un vecteur d'état suivant la terminologie utilisée dans le domaine des systèmes dynamiques.

Notons que l'utilisation des positions facilite l'exploitation d'opérations mathématiques dans l'espace euclidien, cependant à fin d'appliquer des techniques d'apprentissage des contraintes de longueur des segments doivent être ajoutées de manière à bien préserver les distances entre articulations. Au contraire, les rotations, qui s'appliquent sur une structure poly-articulée, présupposent des longueurs de segment fixes ; leur utilisation nécessite de redéfinir les opérations de base dans l'espace $SO3$, en particulier les interpolations, multiplications, normes, inverses, etc. Dans cet espace des rotations, plusieurs descriptions sont possibles, les plus fréquemment utilisées étant les angles d'Euler, les quaternions ou les cartes exponentielles.

Segments ou primitives de mouvement : Dans certaines applications une représentation plus significative du mouvement est nécessaire. Plutôt que de travailler au niveau des postures, il peut être intéressant de segmenter temporellement les exemples de mouvement en *clips* de mouvement plus petits, chaque clip étant considéré comme une action de plus haut niveau, et utilisé pour composer des comportements plus complexes. L'utilisation de segments facilite ainsi la mise en séquence, la recherche et l'analyse des données de mouvement [BSP*04b]. Certains auteurs proposent de regrouper des actions similaires afin de constituer ce qu'ils appellent des *primitives de mouvement* ou des *habilités motrices* [JM02]. En étudiant et modélisant la façon dont ces primitives sont combinées, il est ainsi possible de produire des séquences autonomes.

Représentations à dimension réduite : Pour des mouvements représentés par des postures, l'ensemble des données d'apprentissage peut atteindre de grandes dimensions. Cela s'aggrave lorsque les vitesses et accélérations sont incluses,

la dimension des données étant alors multipliée par trois. En travaillant dans ces espaces de grande dimension, il devient difficile de comprendre et de visualiser toute opération appliquée sur les données. Cela implique également un temps de calcul et des ressources machine accrues.

Dans le contexte des mouvements humains, restreints par les contraintes physiques du squelette rigide et par les contraintes temporelles de l'action en cours, il est alors possible de considérer les données mouvement comme une représentation en haute dimension d'une variété (*manifold* de basse dimension \mathcal{M} plongée dans un espace de haute dimension. À titre d'exemple, la marche humaine est une variété unidimensionnelle qui peut être incorporée dans l'espace des postures [EL13].

Cette hypothèse permet de bénéficier des propriétés topologiques de la variété, à savoir la régularité, la compacité, la connectivité, l'existence d'une équivalence locale à un espace euclidien \mathbb{R}^d et l'existence pour tout voisinage d'un point quelconque v d'une fonction continue bijective \mathcal{F} associant à ce voisinage un sous-espace euclidien \mathbb{R}^d .

Avec un tel formalisme, il est possible d'estimer la fonction \mathcal{F}

$$\mathcal{F} : \mathbb{R}^D \rightarrow \mathbb{R}^d, \text{ with } d \ll D \quad (2)$$

caractérisant le *mapping* entre l'espace de haute dimension des données mouvement dans \mathbb{R}^D et l'espace de basse dimension représentant la variété dans \mathbb{R}^d . Notons que, bien qu'en théorie \mathcal{F} soit une fonction inversible, en pratique les méthodes de réduction de dimension peuvent inférer les informations dans l'espace de basse dimension sans calculer explicitement \mathcal{F} .

La projection des données dans cet espace de basse dimension permet de développer des algorithmes plus efficaces, requérant moins de ressources informatiques, et surtout permet d'appliquer des techniques d'apprentissage et des transformations qui ont un effet équivalent à celui qui serait obtenu dans l'espace de haute dimension.

Données annotées : Dans leur forme brute, les données mouvement manquent de signification ; en particulier, il est pratiquement impossible de dire si des séquences de mouvement de marche, de course, de saut ou de danse sont celles d'un homme, d'une femme ou d'un enfant à partir de l'étude de quelques postures.

Une façon classique d'ajouter du sens (et de la signification) aux mouvements est de les annoter avec des informations d'ordre sémantique (acteurs, styles, émotions, etc.). Ces annotations véhiculent de l'information qui peuvent ensuite guider l'analyse et la synthèse de nouvelles animations.

L'annotation des données mouvement est habituellement un processus manuel, compte tenu du fait que seul un annotateur humain peut percevoir certaines subtilités qui donnent du sens au mouvement. Cependant, certains travaux récents permettent d'effectuer l'annotation automatique des données à partir de méthodes d'apprentissage.

2.3. Pré-traitement des données mouvement

Déformation temporelle : Afin de caractériser les données, il peut s'avérer fondamental de disposer d'une mesure de similarité entre deux ou plusieurs exécutions de la même action. Or chaque exécution diffère des autres spatialement et temporellement (durée, vitesse, etc.). Une autre source de variabilité dans les données provient des différences morphologiques entre sujets.

Afin de maîtriser cette variabilité, il est possible de définir une fonction d'alignement temporel entre les mouvements, c.-à-d., un chemin définissant la correspondance temporelle d'une posture d'un mouvement donné à une posture d'un autre mouvement. Lorsque les mouvements sont ainsi alignés, la similarité spatiale entre les deux mouvements est simple à calculer. Plusieurs algorithmes d'alignement temporel des mouvements sont traditionnellement utilisés : la déformation dynamique temporelle (DTW) [Sen08], la déformation temporelle itérative (ITW) [PSS02] et les interpolations répétitives de postures clés [HLW*13].

Alignement du noeud racine Il a été montré par [KG*02] que les mouvements sont invariants en translation dans le (plan de masse) repère monde et en rotation autour de l'axe vertical, c.-à-d., quel que soit l'endroit de l'espace où le mouvement est exécuté, il demeure le même mouvement composé de la même séquence de postures.

Afin de préserver cette caractéristique et de rendre plus aisée la comparaison entre postures, il est classique, soit de retirer l'information du noeud racine, (puisque c'est l'articulation de ce noeud qui détermine l'emplacement et l'orientation de la posture,) soit d'appliquer des transformations géométriques afin d'aligner les postures avant de les comparer.

Réduction de dimensionnalité : Comme mentionné dans la section §2.2, une bonne façon de gérer le problème de dimensionnalité est de se placer dans un espace de dimension réduite.

De nombreuses méthodes permettent d'estimer la fonction de *mapping* \mathcal{F} . Le choix de cette fonction dépend du fait que les données appartiennent à un sous-espace euclidien, ou bien à une sous-variété de faible dimension. Dans le premier cas, les méthodes de réduction linéaire sont suffisantes. Dans le second cas, des techniques de réduction non linéaires sont nécessaires [SWH*06].

Parmi les méthodes de réduction linéaires, les plus souvent rencontrées sont : l'Analyse en Composantes Principales (ACP), l'ACP probabiliste (ACPP), l'Analyse en Composantes Indépendantes (ACI) et l'Analyse Factorielle. En général, ces quatre méthodes visent à déterminer (les projections de faible dimension ou) les sous-espaces qui maximisent l'information utile contenue dans les données. Chaque méthode propose une manière différente de choisir la base de vecteurs qui définit un tel sous-espace. Par exemple, l'ACI recherche un sous-espace dans lequel les vecteurs de la base sont le plus indépendant possibles d'un point de vue statistique ; l'analyse factorielle, l'ACP et sa variante ACPP recherchent les vecteurs de la base qui maximisent la variance des données.

Des méthodes de réduction non-linéaires et de faible

dimension peuvent également être utilisées. En général, ces méthodes visent à préserver l'information géométrique des données, c.-à-d. les relations de proximité entre les points voisins et éloignés dans l'espace de grande dimension doivent être conservées dans la sous-variété de faible dimension [ST02]. On distingue :

- **Les méthodes globales :** il s'agit de rechercher une représentation des données de faible dimension de telle sorte que les distances deux-à-deux des points dans l'espace de grande dimension sont préservées dans la sous-variété de faible dimension. Ces méthodes sont globales, car les distances entre points voisins et éloignés sont préservées de manière identique. Les méthodes ISOMAP et MDS (*Multidimensional Scaling*) font partie de cette catégorie.
- **Les méthodes locales :** il s'agit de rechercher une représentation à dimension réduite des données de grande dimension qui préserve au mieux la structure locale et les points de proximité originale [SWH*06] ; ces méthodes sont dites locales. Les méthodes LLE (*Locally Linear Embedding*), SOM (*Self Organizing Maps*) et LE (*Laplacian Eigenmap*) font partie de cette catégorie.

3. Les méthodes d'apprentissage basées données

Les approches par apprentissage automatique, dans le contexte de l'animation basée données, présentent une alternative intéressante aux approches basées purement contrôle (synthèse par postures clés par exemple), en permettant la synthèse de nouveaux mouvements *réalistes* et contextualisés. Elles permettent ainsi d'exploiter ou de ré-exploiter des données d'animation disponibles pour produire de nouvelles animations dans le cadre de nouvelles situations, en exploitant des personnages différents et en modifiant les qualités expressives (style, émotion, etc.) des mouvements.

L'apprentissage de la structure des modèles implicites de génération des données mouvement peut être assimilé d'une manière très générale à un problème d'inférence d'une fonction \mathcal{G} qui, à partir d'observables et d'un certain nombre de paramètres, prédit ou produit une nouvelle séquence. Formellement :

$$Y = \mathcal{G}(X, \Psi, \Theta) + \varepsilon \quad (3)$$

où Y est la nouvelle séquence de mouvement produit, X représente les observables dans les bases de données mouvement, Ψ est un ensemble de contraintes limitant le champ des possibles pour les mouvements générés, Θ est un vecteur de paramètres associé à la fonction \mathcal{G} , et ε est un résidu, en général un bruit (identiquement et indépendamment distribué). En fonction de la méthode d'apprentissage et de son cadre d'application, la fonction \mathcal{G} représentera soit une fonction *mapping*, soit une fonction de probabilité, une fonction de classification ou de régression, etc.

Dans cette partie, nous effectuons une revue synthétique des principales méthodes exploitées pour modéliser \mathcal{G} . Pour faciliter la présentation, les méthodes sont réparties en deux catégories : les méthodes statiques et dynamiques.

3.1. Modèles statiques

Les méthodes statiques sont des méthodes qui exploitent uniquement l'état courant de l'animation (positions et expression du personnage animé) pour produire un nouvel état, sans tenir compte des états passés.

3.1.1. Les modèles linéaires à variables latentes

Dans le domaine de l'animation, les méthodes linéaires de réduction de dimension sont exploitées non seulement lors des phases de pré-traitement pour réduire la dimension des observables, mais également lors de la phase de synthèse. Les modèles linéaires à variables latentes, très utilisés en animation [UGB*04], permettent ainsi de représenter les mouvements sous la forme d'une combinaison linéaire de vecteurs de base, également appelés variables latentes.

L'Analyse en Composantes Principales détermine un ensemble orthogonal E de vecteurs de base linéaires[§] e_j (aussi appelés composantes principales) et les coefficients scalaires associés α_j (appelés aussi coefficients principaux), de telle sorte que la variance des données projetées dans le sous-espace défini par les composantes principales soit maximale [Mur12]. Ainsi, toute donnée multidimensionnelle peut-être représentée sous la forme :

$$X_i = \bar{X} + \sum_{j=1}^m \alpha_j e_j \quad (4)$$

où \bar{X} représente la moyenne des données. En interpolant, en mélangeant ou en concaténant les coefficients principaux caractérisant deux mouvements ou plus, de nouveaux mouvements peuvent être synthétisés [UGB*04]. Le principal intérêt de la PCA en animation basée données provient du fait que les mouvements ainsi synthétisés préservent en principe les nuances et les caractéristiques présentes dans les données exemples. Cela s'explique parce que les composantes principales à partir desquelles les synthèses sont produites encapsulent en moyenne les caractéristiques spatio-temporelles et stylistiques des mouvements capturés [RB09]. [CBT07], [EMMT04] et [GBT04] ont exploité avec succès cette approche pour synthétiser de nouveaux mouvements qui respectent des contraintes utilisateur.

L'Analyse en Composantes Indépendantes, de même que l'ACP, permet de séparer différents constituants des mouvements en variables latentes. Ici, les variables latentes sont statistiquement indépendantes, c.-à-d. l'occurrence d'une des variables n'a aucun impact sur l'occurrence d'une autre. Alors que l'ACP maximise la variance des données d'exemple et permet de synthétiser de nouveaux mouvements en modifiant les coefficients principaux, α_j , l'ACI minimise la dépendance entre les variables latentes et permet de synthétiser de nouveaux mouvements en travaillant directement sur ces variables.

Formellement, l'ACI calcule une matrice dite de mélange W pour séparer les vecteurs de base des vecteurs représentant les données d'exemple X :

$$U = W^T X, \text{ with } U = \{u_1, \dots, u_k\} \quad (5)$$

[§]. Les vecteurs de base et les variables latentes sont utilisés comme des expressions synonymes

[SCF06], [MH02], [LPL08] and [KN12] font l'hypothèse que les variables latentes u_i déterminées par l'ACI correspondent aux composantes de style perceptivement significatives présentes dans les mouvements ; en échangeant, fusionnant, raccordant ces composantes, des mouvements présentant des variations de style peuvent être produits. Formellement, un nouveau mouvement Y est produit de la manière suivante :

$$Y = \bar{X} + W^{-1}(C) \quad (6)$$

où \bar{X} est la moyenne calculée sur les données d'exemple, et C représente les composantes de style $\{u_1, \dots, u_k\}$ ajoutées au nouveau mouvement produit.

L'ACP et l'ACI correspondent à une fonction \mathcal{G} de type *mapping*, pour lesquelles X sont les exemples de mouvement, Ψ sont les vecteurs de base e_j et u_k respectivement, et les paramètres Θ sont les coefficients principaux α_i pour l'ACP et la matrice de mélange W pour l'ACI.

3.1.2. Les modèles non-linéaires à variables latentes

Comme précédemment évoqué dans la section 2.3, dans l'hypothèse où les données mouvement décrivent une variété (*manifold*) de basse dimension, des techniques de réduction de dimension non-linéaires (RDNL) peuvent être exploitées pour produire des représentations des données basse dimension. Cependant, en général, les méthodes RDNL ne fournissent pas de modèle explicite pour la fonction d'association (*mapping*) \mathcal{F} . Par conséquent, lorsque l'association est effectuée, aucune méthode simple n'existe pour proposer l'association d'une nouvelle donnée mouvement à une représentation en basse dimension et vice-versa (\mathcal{F}^{-1} n'étant pas non plus explicite) [WFH08].

Une solution possible consiste à apprendre la fonction d'association une fois que la réduction de dimension a été effectuée. Des techniques de régression non-linéaires ou l'inférence de distributions de probabilité liant l'espace d'origine et l'espace des représentations en basse dimension peuvent être tentées pour approcher \mathcal{F} .

Les modèles à base de Processus Gaussien à Variables Latentes (MPGVL) ont été proposés par [Law04] et très largement exploités en animation dans le cadre d'applications diverses. Les PGVL associent un espace de données de grande dimension Y , à un espace dit latent de basse dimension Z , cette association étant assurée par des processus Gaussiens [QDLM08]. Ainsi, la fonction \mathcal{G} est assimilable à la distribution de probabilité conditionnelle suivante :

$$\hat{Y} = \arg \max_Y p(Y|Z, \beta) \quad (7)$$

où Y est un vecteur décrivant une posture q_i , Z représente le vecteur des variables latentes, et β le vecteur des paramètres du processus Gaussien assurant l'association entre X et Y , c'est-à-dire sa moyenne μ et sa matrice de covariance Σ .

Dans une première étape, on estime β en maximisant $p(Y|\beta) = \int p(Y|Z|\beta) dZ$ (en utilisant une variante de l'algorithme EM). La seconde étape consiste à maximiser $\ln p(Z, \beta|Y)$ respectivement à Z en utilisant une méthode de gradient conjugué.

De nombreuses variantes ont été proposées pour améliorer le modèle en fonction des besoins des applications ciblées. [GMHP04] par exemple propose une version de MPGVL intégrant une matrice de mise à l'échelle ; cette matrice détermine les niveaux de variance pour chaque dimension dans l'espace des données mouvement. Pour produire des animations variées et interactives, [LWH*12] suggère d'améliorer l'association des données mouvement avec les variables latentes via une nouvelle distribution d'association a priori. [YAH10] ont mis en oeuvre des MPGVL dits partagés pour animer des personnages non humains à l'aide de mouvements humains ; cette extension permet à des espaces de haute dimension variés de partager un espace latent commun. Dans [UFG*08] et [UFL07] des contraintes topologiques explicites sont appliquées aux MPGVL pour imposer que les données proches dans leur espace d'origine de haute dimension sont bien associées avec des représentations proches dans l'espace latent. Enfin, [WFH07] étendent le principe d'association par processus Gaussien en ajoutant des paramètres d'identité et de style dans l'espace latent de basse dimension.

3.1.3. Les modèles multilinéaires

Les mouvements humains peuvent être vus comme des combinaisons d'éléments multiples très corrélés, caractérisant le contenu, le style ou encore l'identité. Les auteurs définissent par contenu l'action réalisée (par exemple la marche), par style la manière d'exécuter l'action (par exemple une marche lente ou rapide), et par identité les caractéristiques propres à l'acteur (personnalité, genre, morphologie, etc.).

D'un point de vue cinématique, pour produire des animations *stylisées*, il est nécessaire d'analyser et de comprendre les variations de position ou d'orientation angulaire des articulations sollicitées qui encodent ces notions de style et d'expressivité dans les mouvements. A cette fin, certains travaux font état de tentatives pour extraire et séparer le contenu, le style et l'identité des données de mouvement. Des paramètres de contrôle sont ensuite estimés pour chacune de ces trois caractéristiques.

Pour mener à bien cette séparation, [MLC10] ont proposé une approche multilinéaire (analyse n-mode) pour *interpréter* toutes les variations dans les données à partir des trois caractéristiques séparées. L'idée principale repose sur une analyse tensorielle des mouvements, c'est-à-dire une analyse multilinéaire portant sur un ensemble d'espaces vectoriels. La décomposition en valeurs singulières à l'ordre n (n-DVS) est ensuite exploitée pour décomposer ce tenseur en éléments respectivement apparentés au contenu, au style et à l'identité [MK07].

Afin de séparer le contenu des données de mouvement, une représentation en basse dimension de la donnée alignée avec une référence de contenu est calculée [EL04]. Puis, pour séparer le style et l'identité, une approche n-DVS est appliquée sur la fonction d'association qui associe les données mouvement d'origine à leur représentation en basse dimension. Ce faisant, un modèle s'apparentant à notre fonction \mathcal{G} est entraîné [HLW*13] :

$$Y = C \cdot (T \times_2 a^T \times_3 s^T) \quad (8)$$

Pour ce modèle C représente le contenu, T est le tenseur d'ordre 3, a^T est le vecteur caractérisant l'identité de l'acteur, et s^T est le vecteur caractérisant le style.

Pour produire de nouveaux mouvements, des opérations de substitution, de mélange, de fusion et d'interpolation portant sur une, deux ou trois composantes simultanément en fonction de contraintes de haut niveau ou de paramètres permettant de contrôler l'animation.

3.1.4. Méthodes de régression et d'estimation de densité

La génération d'une nouvelle posture ou d'un nouveau mouvement peut-être formulée sous la forme d'une régression qui consiste à prédire la valeur d'une ou plusieurs variables continues y étant donné un vecteur d'entrée x . En général, la fonction \mathcal{G} et par conséquent y sont approchés par une combinaison linéaire de fonctions de base paramétrées $\phi_j(x, \Theta)$:

$$y = \mathcal{G}(x) + \epsilon = \sum_{j=1}^J w_j \phi_j(x, \Theta) + \epsilon \quad (9)$$

Dans le cas le plus simple, les fonctions de base $\phi_j(x, \Theta)$ correspondent aux fonctions linéaires de coefficients Θ . Cependant, il est également possible d'utiliser des bases de fonctions non-linéaires comme c'est le cas par exemple pour les modèles à base de processus Gaussien [CC10] ou encore les méthodes de krigeage (estimation linéaire) [MK05].

Dans les deux cas, x peut-être la représentation en basse dimension d'une posture ou un vecteur de caractéristiques calculés à partir des données d'exemple ou encore des données de synthèse résultant d'un processus d'édition manuelle [IAF09].

Lorsque de nouveaux mouvements sont prédits en exploitant une approche régressive, il est possible de combiner le processus de génération avec un cadre catégoriel, ceci afin de contraindre la synthèse à respecter des paramètres de contrôle de haut niveau caractérisant par exemple un style particulier [CL06], [THB06].

Il est bien évidemment possible d'adopter une perspective probabiliste pour estimer la fonction de régression comme proposé dans [WLZ06a] and [OB08]. Ici, \mathcal{G} est assimilable à une densité de probabilité conditionnelle de produire la sortie y étant donnée l'entrée x et les contraintes ψ :

$$\mathcal{G}(x, y, \psi) = \arg \max_y p(y|x, \psi) \quad (10)$$

3.2. Modèles dynamiques

Les approches statiques pré-supposent que, dans les données mouvement, chaque posture est indépendante et identiquement distribuée. Cependant, un mouvement est intrinsèquement une séquence de postures produite par un système dynamique induisant une dépendance temporelle entre celles-ci. Pour caractériser cet aspect dynamique des données mouvement et la nécessité d'une prise en compte d'un effet mémoire pour améliorer la qualité des synthèses produites, plusieurs méthodes basées apprentissage ont été proposées.

3.2.1. Modèle à base de réseaux bayésiens dynamiques

Les modèles de Markov Cachés (MMC) sont les modèles les plus anciens et les plus utilisés pour modéliser les données séquentielles. Un MMC (Figure 1) est une machine stochastique constituée d'états cachés $\{a_1, \dots, a_3\}$ émettant à chaque transition d'état une observable en sortie $\{O_1, \dots, O_3\}$. Chaque état est caractérisé par une distribution de probabilité d'émission d'observable $P(O_{ij}) = P(O_i|a_j)$ ainsi que par une distribution de probabilité dite de transition permettant de transiter d'un état vers un autre $P(A_{ij}) = P(a_j|a_i)$. A chaque pas de temps, le modèle subit un changement d'état conformément à la distribution $P(A_{ij})$ et produit en sortie une observable conformément à la distribution $P(O_{ij})$. Comme indiqué dans la Figure 1 on peut noter que la probabilité de transiter vers un nouvel état a_j dépend uniquement de l'état précédemment visité a_i , et la probabilité de produire l'observable O_j ne dépend que de l'état courant a_j . En général, comme suggéré dans [EA09], les distributions de probabilité d'émission et de transition sont approchées par des distributions Gaussiennes.

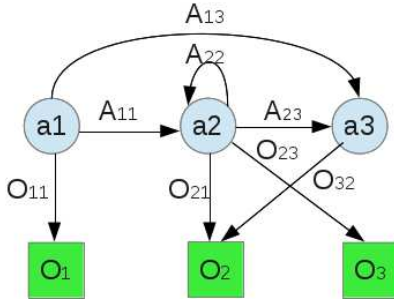


Figure 1: Structure d'un modèle de Markov Caché.

Lorsque les MMC sont exploités en génération de mouvement, quelques modifications vis-à-vis du modèle précédent sont en général apportées. Par exemple, les représentations à la fois des variables cachées et observables peuvent être exprimées dans un espace de basse dimension, et la distribution $P(O_{ij})$ peut-être apprise en exploitant un processus de régression à partir de fonctions Gaussiennes [MHM12]. En exploitant une connaissance d'expert sur les mouvements et leurs variations à produire, [PT09] ont envisagé la construction de MMC hiérarchiques capables de produire des mouvements stylisés. Dans [TMD12], une distribution de probabilité $P(d_i)$ correspondant au temps d_i passé dans l'état a_i peut-être ajoutée dans le modèle. Finalement, il est également possible de contrôler la manière dont les observations sont générées via un vecteur caractérisant le style [BH00] et [WLZ06b].

Les méthodes basées sur les machines de Boltzmann restreintes (MBR) D'une manière similaire aux MMC, les machines de Boltzmann restreintes sont composées d'une structure à deux couches : la couche supérieure contient un ensemble de variables cachées (neurones) tandis que la couche inférieure contient un ensemble de variables visibles. Mais contrairement aux MMC ou autres modèles de type réseau Bayésien, dans un MBR, la couche inférieure est entièrement connectée aux noeuds de la couche supérieure, tan-

dis qu'aucune connexion n'existe à l'intérieur d'une même couche. D'autre part, les connexions entre les couches sont non orientées.

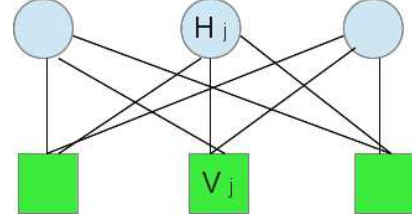


Figure 2: Structure d'une Machine de Boltzmann Restreinte.

Un MBR (Figure 2) est capable de produire de nouvelles observations en activant les états cachés et visibles conformément à une fonction d'énergie qui dépend des poids définis sur les connexions (poids synaptiques) inter-couches. La fonction d'énergie, mais aussi les poids des connexions peuvent être appris grâce à un algorithme d'apprentissage appelé divergence contrastive [THR06]. Cela permet aux MBR de bénéficier d'un mécanisme d'inférence efficace et exact.

En dépit de de ces propriétés, il est clair que de part leur structure, les MBR n'encodent pas les dépendances temporelles et dynamiques. Cependant, en considérant les variables visibles précédemment produites comme des entrées additionnelles et en ajoutant des connexions pour lier ces observations précédentes aux variables visibles courantes ainsi qu'aux variables cachées, [THR11] ont réussi à adjoindre une composante dynamique à la structure d'un MBR classique. Ce nouveau modèle (Figure 3), appelé MBR Conditionnel (MBRC), bénéficie également des mécanismes d'inférence exacts et efficaces. En intégrant un effet mémoire caractéristique des méthodes dynamiques, ce modèle offre une bonne capacité de généralisation pour l'inférence de données séquentielles.

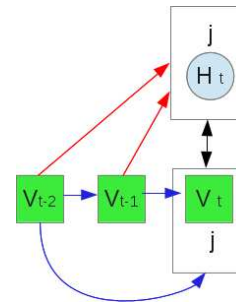


Figure 3: Structure d'un Machine de Boltzmann Restreinte Conditionnelle.

Grâce à cette extension, les nouvelles variables visibles O_t peuvent être produites conditionnellement aux variables cachées H et aux k observations visibles précédentes $(O_{t-1} \dots O_{t-k})$. Formellement, la fonction \mathcal{G} introduite au paragraphe 3 est assimilable à la distribution de probabilité suivante :

$$P(O_t|H, O_{t-1} \dots O_{t-k}) \quad (11)$$

Il est également possible d'ajouter des variables de contrôle ou des contraintes pour réguler les interactions entre les variables visibles (courantes ou précédentes) ou cachées [TH09]. Ainsi, les MBRC peuvent aisément apprendre les transitions entre différents mouvements ou d'un style vers un autre style. Avec cette dernière extension, la fonction \mathcal{G} devient assimilable à $P(O_t|H_t, O_{t-1} \dots O_{t-k}, f_t)$, où f_t représente les paramètres de contrôle.

3.2.2. Représentations compactes et chaînes de Markov

Étant donnée une représentation compacte des exemples de mouvements, par exemple des agrégations (*cluster*) de posture et des vecteurs de caractéristiques en basse dimension (définis manuellement ou automatiquement), il est possible d'approcher l'aspect dynamique des données mouvement en développant des modèles Markoviens.

Pour un modèle de Markov d'ordre K , la posture courante à produire dépend des K postures précédemment générées. Dans le cas de la génération de mouvement, on rend en général la génération d'une nouvelle posture dépendante des exemples de mouvement disponibles dans une base de données. On préserve ainsi la cohérence entre les exemples et les prédictions. Pour la phase d'apprentissage deux types d'approches se distinguent : l'approche globale et l'approche locale.

Pour l'approche globale (en temps différé), les mouvements exemples sont regroupés en plusieurs *clusters* de référence en exploitant une mesure de similarité appropriée. Une mesure de probabilité de transition entre postures [LCR*02] ou entre *clusters* [TH00] est alors estimée. Il est également possible d'ajouter à chaque *cluster* un modèle pour contrôler les variations de styles observables sur les postures qui lui sont rattachées [MC12]. Pour produire un nouveau mouvement, un chemin optimal entre les clusters est recherché afin de respecter au mieux les contraintes de haut niveau imposées. Les postures successives sont alors sélectionnées le long du chemin optimal au sein des clusters.

Pour l'approche locale (*temps réel*), une nouvelle posture y_t est produite en tenant compte des K postures précédemment générées $\{\tilde{y}_{t-k}|k = 1, \dots, K\}$ et des J extraits de mouvement $\{x_j|j = 1, \dots, J\}$ les plus proches de $\tilde{y}_{t-1}, \tilde{y}_{t-2} \dots \tilde{y}_{t-k}$ contenus dans la bases d'exemples. Les contraintes c_t quant à elles viennent conditionner la mesure de proximité utilisée pour extraire les $\{x_j\}$. Pour accélérer la recherche des J exemples les plus proches, des techniques de réduction de dimension [LHC*11] et des graphes de voisinage [CH05] sont utilisés.

En exploitant les \tilde{y}_{t-k} et les x_j , la prédiction d'une nouvelle posture peut-être formulée sous la forme d'une régression linéaire :

$$y_t = \Lambda^T Y + \varepsilon \quad (12)$$

où Λ sont les coefficients de régression appris à partir des exemples les plus proches x_j , $Y = \{\tilde{y}_{t-k}|k = 1, \dots, K\}$ sont les postures précédentes, et ε est un vecteur de bruit [LWC*11].

3.2.3. Modèles à base de système dynamiques

Les systèmes dynamiques sont utilisés pour modéliser des phénomènes physiques, dont les variables observables y_t changent au cours du temps conformément aux changements d'états internes x_t du système. Dans le cas de l'animation de personnage, le phénomène à modéliser est le mouvement humain et les variables observables sont représentées par les postures, les vitesses et les accélérations. Les systèmes dynamiques représentant des phénomènes physiques sont caractérisés en général par des états continus [Mur12]. Les modèles de Markov Cachés sont des cas particuliers de systèmes dynamiques pour lesquels, en général, les états sont discrets.

Dans le contexte de l'animation, une classe de modèles dynamiques est particulièrement exploitée : il s'agit des systèmes linéaires invariants dans le temps (SLIT). Pour de tels systèmes, si à un instant t l'état caché x_t produit une sortie y_t , ce même état produira la même sortie à tout instant, par exemple à $t + k$. Formellement, un SLIT peut-être défini comme suit :

$$\begin{aligned} x_{t+1} &= Ax_t + V \\ y_t &= Cx_t + W \end{aligned} \quad (13)$$

où A et C sont respectivement les matrices de transition et d'observation, V et W sont des bruits Gaussiens indépendants. Tous ces paramètres peuvent être appris en exploitant des techniques d'identification ou estimés à partir d'exemples [HPP05].

En général, les états x_t sont cachés et peuvent être vus comme des représentations en basse dimension des exemples de mouvement [CH07], [WFH08] and [ZZW08].

Une extension de SLIT a également été proposée pour produire de nouveaux mouvements. Au lieu de travailler sur la base d'une représentation des postures, des primitives de mouvement sont définies et modélisées par des SLIT élémentaires [LWS02]. Puis, en exploitant une chaîne de Markov [PRM00], il devient possible de synthétiser des mouvements complexes tels que la danse en transitant d'un modèle SLIT élémentaire à un autre.

4. Discussion

L'utilisation conjointe de données de mouvement capturées et de méthodes d'apprentissage automatique est une tendance dominante dans le domaine de l'animation de personnages. Cependant à notre connaissance il n'existe aucun protocole d'évaluation qui nous permettrait de statuer sur les méthodes qui marchent le mieux dans un contexte donné. Pour cette raison, nous revenons sur les inconvénients mis en évidence pour chacune des méthodes décrites précédemment ainsi que sur les solutions avancées pour ces problèmes, à la fois dans la communauté de l'apprentissage automatique et de l'animation par ordinateur.

Malgré leur utilisation très répandue, les méthodes de réduction de dimension linéaires font l'hypothèse que le mouvement humain est linéaire et peut être représenté par une combinaison de quelques composantes principales. Cette représentation en basse dimension marche bien pour un ensemble d'exemples composé de mouvements courts, simples

et similaires comportant peu de variabilité. Cependant, plus les mouvements sont diversifiés, plus le nombre de composantes doit être important pour représenter toute la variabilité des données [BSP*04a]. Dans ce cas, on perd l'intérêt de travailler dans un espace de basse dimension. De plus, il est montré que les données de mouvement sont par nature hautement non linéaires, ce qui conduit à la difficulté de les appréhender par des méthodes de réduction linéaires [EL04].

En suivant ce raisonnement, l'utilisation de méthodes de réduction non-linéaires est pleinement justifié. Des techniques telles que ISOMAP et LLE donnent une représentation basse dimension du mouvement tout en préservant les non linéarités intrinsèques du mouvement humain [WFH08]. Néanmoins, la plupart de ces méthodes ne possèdent pas de fonction d'association entre les espaces de haute dimension et de basse dimension. Dans ce contexte, les méthodes offrant un *mapping* direct et un espace de représentation latent non linéaire deviennent plus attractives.

La méthode MGPLV offre ces deux fonctionnalités, ce qui explique pourquoi elle a été largement utilisée ces dernières années. Elle dispose d'une fonction de probabilité qui apprend l'association entre les variables latentes et les vecteurs de caractéristiques qui peut être utilisée par exemple pour résoudre un problème de cinématique inverse [GMHP04]. Dans sa forme standard, la méthode MGPLV ne garantit pas que la structure géométrique et topologique d'un mouvement soit correctement préservée dans l'espace latent, ce qui revient à dire que les postures les plus proches ne sont pas nécessairement celles qui correspondent aux points les plus proches dans l'espace latent [UFG*08]. Pour contourner cette limitation, des contraintes de connectivité antérieures et des contraintes topologiques ont été ajoutées au modèle standard [LWH*12], [UFL07]. En outre, lorsque des styles multiples, des variations morphologiques ou d'activité sont présents dans les données, les méthodes basées GPLVM conduisent à des animations non fluides et ne parviennent pas à généraliser [WFH07].

Pour des mouvements stylistiquement différents, les modèles bilinéaires et multilinéaires sont très fréquemment utilisés, car ils constituent un modèle génératif unique couvrant tous les exemples de la base de données [MLC10]. Ces méthodes sont basées sur l'idée que les différentes exécutions d'une même action partagent un contenu commun dit "neutre", le style étant ramené à des variations de position ou de rotation ajoutées à cette action neutre. En dépit des bons résultats, il n'existe pas d'étude ou d'évaluation perceptive qui justifie de découper l'espace en trois axes orthogonaux, l'un représentant le style, l'autre le contenu et parfois un troisième l'identité. En outre, ces modèles ne traitent pas des transitions entre les activités ou les styles [MLC10].

Les méthodes de réduction de dimension, de même que les modèles MGPLV et multilinéaires considèrent uniquement les aspects spatiaux et temporels instantanés (vitesse et accélération) des données de mouvement. Ils ignorent la nature séquentielle et conditionnelle des données et par conséquent n'exploitent pas complètement la nature dynamique des exemples [MP08]. Pour prendre en compte cet aspect dynamique, de simples modèles tels que les MMC et les SLIT ont été proposés. Ces deux derniers modèles exploitent la

notion d'états cachés et produisent des mouvements suivant une fonction ou une probabilité de transition état-à-état, ainsi qu'une fonction ou une probabilité d'émission.

Bien que ces deux méthodes, MMC et SLIT, produisent de bons résultats, leur capacité à intégrer des caractéristiques d'expressivité pour des mouvements plus complexes est limitée [WFH07]. De leur côté, les modèles SLIT ne gèrent pas bien la dynamique non linéaire présente dans le mouvement humain, car chaque observation est issue d'un système linéaire [THR06]. Des améliorations telles que le modèle *Switching SLIT* et l'introduction d'une hiérarchie qui s'applique aux états cachés d'un MMC ont été proposées pour accroître l'expressivité des modèles précédents. Cependant ces extensions nécessitent de grands volumes de données d'entraînement et requièrent beaucoup de paramètres qui doivent être réglés à la main [UFL07].

Malgré les limitations des modèles SLIT et MMC, ils peuvent être utilisés comme structure de base de modèles plus puissants tels que les représentations compactes ayant des propriétés Markoviennes. Ce nouveau type de méthodes conduit à des résultats prometteurs pour l'inférence de postures. En considérant uniquement des exemples qui sont proches des postures précédentes synthétisées et de certaines contraintes, ils évitent la nature exponentielle des MMC ainsi que la limitation relative à la linéarité des modèles SLIT. Cette spécificité rend possible l'approximation d'un modèle dynamique de mouvement humain à travers des modèles dynamiques locaux capables de prédire un large spectre de nouvelles postures [LWC*11]. Néanmoins ils requièrent des exemples d'apprentissage pendant la phase d'estimation et leurs résultats relèvent hautement de l'existence d'exemples proches de l'état courant (contraintes et postures précédentes) du système. Si aucun ou peu d'exemples sont trouvés, les prédictions résultantes sont moins précises [CH07]. De plus des méthodes d'indexation appropriées sont requises [KTWZ10].

Les machines de Boltzmann restreintes conditionnelles MBRC ont fait leurs preuves pour surmonter bon nombre des inconvénients énumérés ci-dessus. Ils approchent les performances des modèles dynamiques globaux via l'introduction de poids sur les arêtes et d'une fonction d'activation ; les exemples n'ont pas besoin d'être conservés après la formation du modèle, ils peuvent être facilement adaptés à des masses de données ; enfin, comme les états cachés sont distribués (le modèle est entièrement connecté), ils atteignent la puissance expressive des modèles MMC et LIST [TH09]. Cependant, ces modèles sont difficiles à apprendre - une seule des méthodes d'apprentissage est connue - et à contrôler [WFH07]. En outre, ils doivent disposer de plusieurs postures initiales pour prédire un nouveau mouvement ; enfin aucune implémentation en temps réel n'a été réalisée [MHM12]. Par ailleurs, il est nécessaire de disposer d'exemples en nombre suffisant lors de l'apprentissage, afin de faire la synthèse complète de nouveaux mouvements ou de nouveaux styles [CM11].

Deux inconvénients majeurs de toutes les méthodes discutées peuvent être relevés : i) tout d'abord, elles se focalisent sur un petit ensemble de mouvements, dans la plupart des cas des exemples de locomotion très similaires entre

Méthode	Re-utilisation en synthèse	Re-utilisation en style	Re-utilisation en édition
ACP	*	*	
ICA		*	
MGPLV et variations	*	*	*
Modèles multilinéaires	*	*	
Régression et estimation de densité	*		*
Modèles de Markov cachés	*	*	
Machines de Boltzmann restreintes	*	*	
Représentations compactes et chaînes de Markov	*	*	*
Modèles à base de système dynamiques	*	*	

Table 1: Méthodes et applications

eux. Il y a peu d'études qui montrent comment ces méthodes marchent sur des mouvements non cycliques, plus complexes et moins pré-traités. ii) De plus, la généralisation à de nouveaux mouvements et styles est limitée par la quantité d'exemples disponibles pendant l'apprentissage. Comme il est presque impossible de capturer tous les mouvements humains et toutes leurs variations possibles, la plupart des méthodes restreignent les exemples et les estimations de façon qu'ils soient aussi proches que possible les uns des autres, ce qui permet d'éviter le sur-ajustement [CM11].

Enfin, nous tenons à souligner que très peu de méthodes font usage de l'information sémantique pour encoder les données de mouvement. Nous sommes convaincus que beaucoup de méthodes d'analyse / synthèse pourraient dépasser certaines de leurs limitations en utilisant ces informations de haut niveau fournies par des experts ou des utilisateurs.

5. Applications

Les méthodes d'apprentissage mentionnées précédemment ont été utilisées dans le cadre d'un large éventail d'applications, non seulement pour l'animation de personnages, mais aussi pour le contrôle d'avatar et le suivi de mouvement en vision par ordinateur. Dans cette section, nous listons quelques unes d'entre elles.

Les méthodes de réduction à la fois linéaires et non linéaires sont exploitées pour résoudre des problèmes de cinématique inverse (estimation des angles aux articulations à partir des positions des effecteurs extrémité) [OH06], [QYWW11] and [LXW07]. Elles sont également appliquées pour le contrôle d'avatars et le suivi de mouvements à partir de données d'entrée provenant de quelques capteurs [LZWM06], [XFCQ08].

Les méthodes multilinéaires [MCC09], [LPL09] ainsi que les méthodes basées système dynamiques sont utilisées avec succès pour la production de mouvements stylisés [BSS03]. Ces dernières méthodes ont également été utilisées pour reconstruire des postures humaines à partir de silhouettes [ETL08] et de capteurs à bas coûts [XKC*08].

Les méthodes basées régression, en particulier les régressions à base de processus Gaussiens, ont permis de produire avec succès de nouvelles animations interactives et stylisées, p.ex. dans le cadre de la locomotion [MXH*10], [WX11]. Elles ont également apporté des contributions significatives au problème de l'édition de mouvement et

d'adaptation morphologique [IAF09], [YAH10]. Elles fournissent également des résultats prometteurs dans le cadre de la génération d'animations dites "physiquement correctes" [WMC11], [YL10].

Une vue d'ensemble des méthodes et de leurs applications est présentée dans le tableau 1.

6. Conclusion

Dans cette étude nous avons présenté une synthèse des méthodes basées apprentissage qui ont été utilisées ces 13 dernières années dans le cadre de l'animation de personnage. Nous avons décrit les principes sous-tendant chaque méthode ainsi que leurs principales caractéristiques et inconvénients. Nous avons également classé ces méthodes suivant la nature séquentielle et conditionnelle des données de mouvement humain.

Les travaux de recherche futurs dans le domaine de l'animation de personnages virtuels devront aborder de nombreuses questions qui restent ouvertes, afin de résoudre le problème de la génération de nouveaux mouvements à partir d'exemples. Ces questions concernent l'utilisation des annotations sémantiques, la généralisation des mouvements au-delà de l'espace défini par l'ensemble d'apprentissage, une analyse perceptuelle approfondie des notions de contenu, style et identité, et l'intégration de la connaissance du domaine pour améliorer l'apprentissage.

Indépendamment des limitations rencontrées dans les approches d'apprentissage actuelles, ces approches sont toujours considérées comme une solution attrayante pour le problème de l'animation et resteront un sujet de recherche important.

Références

- [AF02] ARIKAN O., FORSYTH D. A. : Interactive motion generation from examples. *ACM Transactions on Graphics (TOG)*. Vol. 21, Num. 3 (2002), 483–490.
- [AFO03] ARIKAN O., FORSYTH D. A., O'BRIEN J. F. : Motion synthesis from annotations. *ACM Transactions on Graphics (TOG)*. Vol. 22, Num. 3 (2003), 402–408.
- [AI06] ARIKAN O., IKEMOTO L. : *Computational Studies of Human Motion : Tracking and Motion Synthesis*. Now Publishers Inc, 2006.

- [BH00] BRAND M., HERTZMANN A. : Style machines. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (2000), SIGGRAPH '00, pp. 183–192.
- [BSP*04a] BARBIĆ J., SAFONOVA A., PAN J.-Y., FALOUTSOS C., HODGINS J. K., POLLARD N. S. : Segmenting motion capture data into distinct behaviors. In *Proceedings of Graphics Interface 2004* (2004), Canadian Human-Computer Communications Society, pp. 185–194.
- [BSP*04b] BARBIĆ J., SAFONOVA A., PAN J.-Y., FALOUTSOS C., HODGINS J. K., POLLARD N. S. : Segmenting motion capture data into distinct behaviors. In *Proceedings of Graphics Interface 2004* (2004), pp. 185–194.
- [BSS03] BISSACCO A., SAISAN P., SOATTO S. : Modeling human gaits with subtleties. In *Proceedings of 13th IFAC Symposium on System Identification* (2003).
- [BW95] BRUDERLIN A., WILLIAMS L. : Motion signal processing. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques* (1995), SIGGRAPH '95, pp. 97–104.
- [CBT07] CARVALHO S. R., BOULIC R., THALMANN D. : Interactive low-dimensional human motion synthesis by combining motion models and pik. *Computer Animation and Virtual Worlds*. Vol. 18, Num. 4-5 (2007), 493–503.
- [CC10] COURTY N., CUZOL A. : Conditional stochastic simulation for character animation. *Computer Animation and Virtual Worlds*. Vol. 21, Num. 3-4 (2010), 443–452.
- [CH05] CHAI J., HODGINS J. K. : Performance animation from low-dimensional control signals. In *ACM Transactions on Graphics (TOG)* (2005), vol. 24, ACM, pp. 686–696.
- [CH07] CHAI J., HODGINS J. K. : Constraint-based motion optimization using a statistical dynamic model. *ACM Transactions on Graphics (TOG)*. Vol. 26, Num. 3 (2007), 8.
- [CL06] CHIEN Y.-R., LIU J.-S. : Learning the stylistic similarity between human motions. In *Advances in Visual Computing*, vol. 4291 de *Lecture Notes in Computer Science*. 2006.
- [CM11] CHIU C.-C., MARSELLA S. : A style controller for generating virtual human behaviors. In *AAMAS* (2011), pp. 1023–1030.
- [EA09] ETEMAD S., ARYA A. : Recognition and re-synthesis of 3d human motion with personalized variations. In *Multimedia Computing and Systems, 2009. ICMCS '09. International Conference on* (2009), pp. 106–111.
- [EL04] ELGAMMAL A., LEE C.-S. : Separating style and content on a nonlinear manifold. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on* (2004), vol. 1, IEEE, pp. 1–478.
- [EL13] ELGAMMAL A., LEE C.-S. : Homeomorphic manifold analysis (hma) : Generalized separation of style and content on manifolds. *Image and Vision Computing*. Vol. 31, Num. 4 (2013), 291 – 310.
- [EMMT04] EGGES A., MOLET T., MAGNENAT-THALMANN N. : Personalised real-time idle motion synthesis. In *Computer Graphics and Applications, 2004. PG 2004. Proceedings. 12th Pacific Conference on* (2004), IEEE, pp. 121–130.
- [ETL08] EK C. H., TORR P. H., LAWRENCE N. D. : Gaussian process latent variable models for human pose estimation. In *Machine learning for multimodal interaction*. Springer, 2008, pp. 132–143.
- [FHKS12] FENG A. W., HUANG Y., KALLMANN M., SHAPIRO A. : An analysis of motion blending techniques. In *The Fifth International Conference on Motion in Games* (2012).
- [GBT04] GLARDON P., BOULIC R., THALMANN D. : A coherent locomotion engine extrapolating beyond experimental data.
- [GMHP04] GROCHOW K., MARTIN S. L., HERTZMANN A., POPOVIĆ Z. : Style-based inverse kinematics. *ACM Transactions on Graphics*. Vol. 23, Num. 3 (2004), 522–531.
- [GP12] GEIJTENBEEK T., PRONOST N. : Interactive character animation using simulated physics : A state-of-the-art review. In *Computer Graphics Forum* (2012), vol. 31, Wiley Online Library, pp. 2492–2515.
- [Her03] HERTZMANN A. : Machine learning for computer graphics : A manifesto and tutorial. In *Computer Graphics and Applications, 2003. Proceedings. 11th Pacific Conference on* (2003), IEEE, pp. 22–36.
- [HLW*13] HE Z., LIANG X., WANG J., ZHAO Q., GUO C. : Flexible editing of human motion by three-way decomposition. *Computer Animation and Virtual Worlds* (2013).
- [HPP05] HSU E., PULLI K., POPOVIĆ J. : Style translation for human motion. In *ACM Transactions on Graphics (TOG)* (2005), vol. 24, ACM, pp. 1082–1089.
- [IAF09] IKEMOTO L., ARIKAN O., FORSYTH D. : Generalizing motion edits with gaussian processes. *ACM Transactions on Graphics (TOG)*. Vol. 28, Num. 1 (2009), 1.
- [JM02] JENKINS O. C., MATARIC M. J. : Deriving action and behavior primitives from human motion data. In *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on* (2002), vol. 3, IEEE, pp. 2551–2556.
- [Joh73] JOHANSSON G. : Visual perception of biological motion and a model for its analysis. *Perception & psychophysics*. Vol. 14, Num. 2 (1973), 201–211.
- [KG*02] KOVAR L., GLEICHER M., ET AL. : Motion graphs. *ACM Transactions on Graphics (TOG)*. Vol. 21, Num. 3 (2002), 473–482.
- [KN12] KIM Y., NEFF M. : Component-based locomotion composition. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2012), SCA '12, pp. 165–173.
- [KTWZ10] KRÜGER B., TAUTGES J., WEBER A., ZINKE A. : Fast local and global similarity searches in large motion capture databases. In *2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (juillet 2010), SCA '10, pp. 1–10.
- [Law04] LAWRENCE N. D. : Gaussian process latent variable models for visualisation of high dimensional data. *Advances in neural information processing systems*. Vol. 16, Num. 329-336 (2004), 3.
- [LCR*02] LEE J., CHAI J., REITSMA P. S., HODGINS J. K., POLLARD N. S. : Interactive control of avatars animated with human motion data. In *ACM Transactions on Graphics (TOG)* (2002), vol. 21, ACM, pp. 491–500.
- [LHC*11] LIU H., HE F., CAI X., CHEN X., CHEN Z. : Human motion synthesis using window-based local principal component analysis. In *Computer-Aided Design and Computer Graphics (CAD/Graphics), 2011 12th International Conference on* (2011), pp. 282–287.
- [LPL08] LIU G., PAN Z., LIN Z. : Style subspaces for character animation. *Computer Animation and Virtual Worlds*. Vol. 19, Num. 3-4 (2008), 199–209.

- [LPL09] LIU G., PAN Z., LI L. : Motion synthesis using style-editable inverse kinematics. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents* (2009), IVA '09, pp. 118–124.
- [LWC*11] LIU H., WEI X., CHAI J., HA I., RHEE T. : Realtime human motion control with a small number of inertial sensors. In *Symposium on Interactive 3D Graphics and Games* (2011), ACM, pp. 133–140.
- [LWH*12] LEVINE S., WANG J. M., HARAUX A., POPOVIĆ Z., KOLTUN V. : Continuous character control with low-dimensional embeddings. *ACM Transactions on Graphics*. Vol. 31, Num. 4 (2012), 28.
- [LWS02] LI Y., WANG T., SHUM H.-Y. : Motion texture : a two-level statistical model for character motion synthesis. In *ACM Transactions on Graphics (TOG)* (2002), vol. 21, ACM, pp. 465–472.
- [LXW07] LI C., XIA S., WANG Z. : Pose synthesis using the inverse of jacobian matrix learned from examples. In *Virtual Reality Conference, 2007. VR'07. IEEE* (2007), IEEE, pp. 99–106.
- [LZWM06] LIU G., ZHANG J., WANG W., MCMILLAN L. : Human motion estimation from a reduced marker set. In *Proceedings of the 2006 symposium on Interactive 3D graphics and games* (2006), I3D '06, pp. 35–42.
- [MC12] MIN J., CHAI J. : Motion graphs++ : a compact generative model for semantic motion analysis and synthesis. *ACM Transactions on Graphics (TOG)*. Vol. 31, Num. 6 (2012), 153.
- [MCC09] MIN J., CHEN Y., CHAI J. : Interactive generation of human animation with deformable motion models. *ACM Transactions on Graphics (TOG)*. Vol. 29, Num. 1 (2009), 9.
- [MH02] MORI H., HOSHINO J. : Independent component analysis and synthesis of human motion. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on* (2002), vol. 4, IEEE, pp. IV–3564.
- [MHK06] MOESLUND T. B., HILTON A., KRÜGER V. : A survey of advances in vision-based human motion capture and analysis. *Computer vision and image understanding*. Vol. 104, Num. 2 (2006), 90–126.
- [MHM12] MATSUBARA T., HYON S.-H., MORIMOTO J. : Real-time stylistic prediction for whole-body human motions. *Neural Networks*. Vol. 25 (2012), 191–199.
- [MK05] MUKAI T., KURIYAMA S. : Geostatistical motion interpolation. In *ACM Transactions on Graphics (TOG)* (2005), vol. 24, ACM, pp. 1062–1070.
- [MK07] MUKAI T., KURIYAMA S. : Multilinear motion synthesis with level-of-detail controls. In *Computer Graphics and Applications, 2007. PG'07. 15th Pacific Conference on* (2007), IEEE, pp. 9–17.
- [MLC10] MIN J., LIU H., CHAI J. : Synthesis and editing of personalized stylistic human motion. In *Proceedings of the 2010 ACM SIGGRAPH symposium on Interactive 3D Graphics and Games* (2010), ACM, pp. 39–46.
- [MP08] MOON K., PAVLOVIĆ V. : Graphical models for human motion modelling. In *Human Motion*. Springer, 2008, pp. 159–183.
- [Mur12] MURPHY K. P. : *Machine Learning : A Probabilistic Perspective*. The MIT Press, 2012.
- [MXH*10] MA W., XIA S., HODGINS J. K., YANG X., LI C., WANG Z. : Modeling style and variation in human motion. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2010), Eurographics Association, pp. 21–30.
- [OB08] OKWECHIME D., BOWDEN R. : A generative model for motion synthesis and blending using probability density estimation. In *Articulated Motion and Deformable Objects*. Springer, 2008, pp. 218–227.
- [OH06] ONG E.-J., HILTON A. : Learnt inverse kinematics for animation synthesis. *Graphical Models*. Vol. 68, Num. 5 (2006), 472–483.
- [PP10] PEJSA T., PANDZIC I. S. : State of the art in example-based motion synthesis for virtual characters in interactive applications. In *Computer Graphics Forum* (2010), vol. 29, Wiley Online Library, pp. 202–226.
- [PRM00] PAVLOVIC V., REHG J. M., MACCORMICK J. : Learning switching linear models of human motion. In *NIPS* (2000), Citeseer, pp. 981–987.
- [PSS02] PARK S. I., SHIN H. J., SHIN S. Y. : On-line locomotion generation based on motion blending. In *Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation* (2002), ACM, pp. 105–111.
- [PT09] PAN W., TORRESANI L. : Unsupervised hierarchical modeling of locomotion styles. In *Proceedings of the 26th Annual International Conference on Machine Learning* (2009), ACM, pp. 785–792.
- [QDLM08] QUIRION S., DUCHESNE C., LAURENDEAU D., MARCHAND M. : Comparing gplvm approaches for dimensionality reduction in character animation.
- [QYWW11] QU S., YU R., WEI Y., WU L. : Gaussian process latent variable models for inverse kinematics. *Journal of Multimedia*. Vol. 6, Num. 1 (2011), 48–55.
- [RB09] RAUNHARDT D., BOULIC R. : Motion constraint. *The Visual Computer*. Vol. 25, Num. 5-7 (2009), 509–518.
- [SCF06] SHAPIRO A., CAO Y., FALOUTSOS P. : Style components. In *Proceedings of Graphics Interface 2006* (2006), Canadian Information Processing Society, pp. 33–39.
- [Sen08] SENIN P. : Dynamic time warping algorithm review. *Honolulu, USA* (2008).
- [ST02] SILVA V. D., TENENBAUM J. B. : Global versus local methods in nonlinear dimensionality reduction. In *Advances in neural information processing systems* (2002), pp. 705–712.
- [SWH*06] SAUL L. K., WEINBERGER K. Q., HAM J. H., SHAFER, LEE D. D. : Spectral methods for dimensionality reduction. *Semisupervised learning* (2006), 293–308.
- [TH00] TANCO L. M., HILTON A. : Realistic synthesis of novel human movements from a database of motion capture examples. In *Human Motion, 2000. Proceedings. Workshop on* (2000), IEEE, pp. 137–142.
- [TH09] TAYLOR G. W., HINTON G. E. : Factored conditional restricted boltzmann machines for modeling motion style. In *Proceedings of the 26th annual international conference on machine learning* (2009), ACM, pp. 1025–1032.
- [THB06] TORRESANI L., HACKNEY P., BREGLER C. : Learning motion style synthesis from perceptual observations. In *Advances in Neural Information Processing Systems* (2006), pp. 1393–1400.
- [THR06] TAYLOR G. W., HINTON G. E., ROWEIS S. T. : Modeling human motion using binary latent variables. In *Advances in neural information processing systems* (2006), pp. 1345–1352.
- [THR11] TAYLOR G. W., HINTON G. E., ROWEIS S. T. : citwo distributed-state models for generating high-dimensional time series. *Journal of Machine Learning Research*. Vol. 12 (2011), 1025–1068.

- [TMD12] TILMANNE J., MOINET A., DUTOIT T. : Stylistic gait synthesis based on hidden markov models. *EURASIP Journal on advances in signal processing*. Vol. 2012, Num. 1 (2012), 1–14.
- [UFG*08] URTASUN R., FLEET D. J., GEIGER A., POPOVIĆ J., DARRELL T. J., LAWRENCE N. D. : Topologically-constrained latent variable models. In *Proceedings of the 25th international conference on Machine learning* (2008), ACM, pp. 1080–1087.
- [UFL07] URTASUN R., FLEET D. J., LAWRENCE N. D. : Modeling human locomotion with topologically constrained latent variable models. In *Human Motion—Understanding, Modeling, Capture and Animation*. Springer, 2007, pp. 104–118.
- [UGB*04] URTASUN R., GLARDON P., BOULIC R., THALMANN D., FUA P. : Style-based motion synthesis. In *Computer Graphics Forum* (2004), vol. 23, Wiley Online Library, pp. 799–812.
- [VWVBE*10] VAN WELBERGEN H., VAN BASTEN B. J., EGGES A., RUTTKAY Z. M., OVERMARS M. H. : Real time animation of virtual humans : A trade-off between naturalness and control. In *Computer Graphics Forum* (2010), vol. 29, pp. 2530–2554.
- [WFH07] WANG J. M., FLEET D. J., HERTZMANN A. : Multi-factor gaussian process models for style-content separation. In *Proceedings of the 24th international conference on Machine learning* (2007), ACM, pp. 975–982.
- [WFH08] WANG J. M., FLEET D. J., HERTZMANN A. : Gaussian process dynamical models for human motion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. Vol. 30, Num. 2 (2008), 283–298.
- [WLZ06a] WANG Y., LIU Z.-Q., ZHOU L.-Z. : Key-styling : learning motion style for real-time synthesis of 3d animation. *Computer Animation and Virtual Worlds*. Vol. 17, Num. 3-4 (2006), 229–237.
- [WLZ06b] WANG Y., LIU Z.-Q., ZHOU L.-Z. : Learning style-directed dynamics of human motion for automatic motion synthesis. In *Systems, Man and Cybernetics, 2006. SMC'06. IEEE International Conference on* (2006), vol. 5, IEEE, pp. 4428–4433.
- [WMC11] WEI X., MIN J., CHAI J. : Physically valid statistical models for human motion generation. *ACM Transactions on Graphics (TOG)*. Vol. 30, Num. 3 (2011), 19.
- [WX11] WANG J., XIA S. : Layered interpolation for interactive avatar control. In *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry* (2011), ACM, pp. 49–58.
- [XFCQ08] XIE L., FANG B., CAO Y., QUEK F. : A nonlinear manifold learning framework for real-time motion estimation using low-cost sensors. In *Proceedings of the 2008 37th IEEE Applied Imagery Pattern Recognition Workshop* (2008), AIPR '08, pp. 1–8.
- [XKC*08] XIE L., KUMAR M., CAO Y., GRACANIN D., QUEK F. : Data-driven motion estimation with low-cost sensors. In *Visual Information Engineering, 2008. VIE 2008. 5th International Conference on* (2008), pp. 600–605.
- [YAH10] YAMANE K., ARIKI Y., HODGINS J. : Animating non-humanoid characters with human motion data. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2010), Eurographics Association, pp. 169–178.
- [YL10] YE Y., LIU C. K. : Synthesis of responsive motion using a dynamic model. In *Computer Graphics Forum* (2010), vol. 29, Wiley Online Library, pp. 555–562.
- [ZZW08] ZHANG Y., ZHU D., WANG Z. : Motion synthesis from the semantic signals. *International Journal of Image and Graphics*. Vol. 08, Num. 04 (2008), 535–550.